(1) If you are writing a program to read in data from a file, what statements or options might you need to use if there are missing values in the data?

(2) How might you determine if there are missing values in a data set?

(3) Is missing data always a problem that should be investigated? Why or why not?

(4) When manipulating data that has missing values, is it better to use an assignment statement or a function? Why?

(1) By using the MISSOVER option on the INFILE statement, the missing value is assigned to a variable without skipping over the line.

By using the TRUNCOVER option on the INFILE statement, the missing values will be fill in the unfilled variable.

(2) To determine the presence of missing values in a data set, the programmer can use the missing function, using a SUM statement to count the amount of missing values, or using the PROC MEANS statement with the nmiss option.

(3) The analyzer should be aware of the presence of missing data, whether it be only a few or many points of missing data. It may or may not be a problem depending on the type of analysis, algorithms, or statistics which are to be employed on the dataset.

(4) Potentially both would be fine, depending on the type of analysis to be completed. The advantage of using a function it that it would skip over missing data.